EX:No.2 DATE:1/02/25	**Implement programs for visualizing time series data.**

AIM:

To Implement programs for visualizing time series data.

OBJECTIVE:

To analyze and visualize Water pollution trends from 2012 to 2021 using multiple time-series plots.

BACKGROUND:

- Air pollution impacts health, climate, and the environment.
- Major pollutants include PM2.5, CO, NO2, SO2, and O3.
- Understanding trends helps in regulation and policy-making.
- Data visualization makes patterns easier to interpret.

SCOPE OF THE PROGRAM:

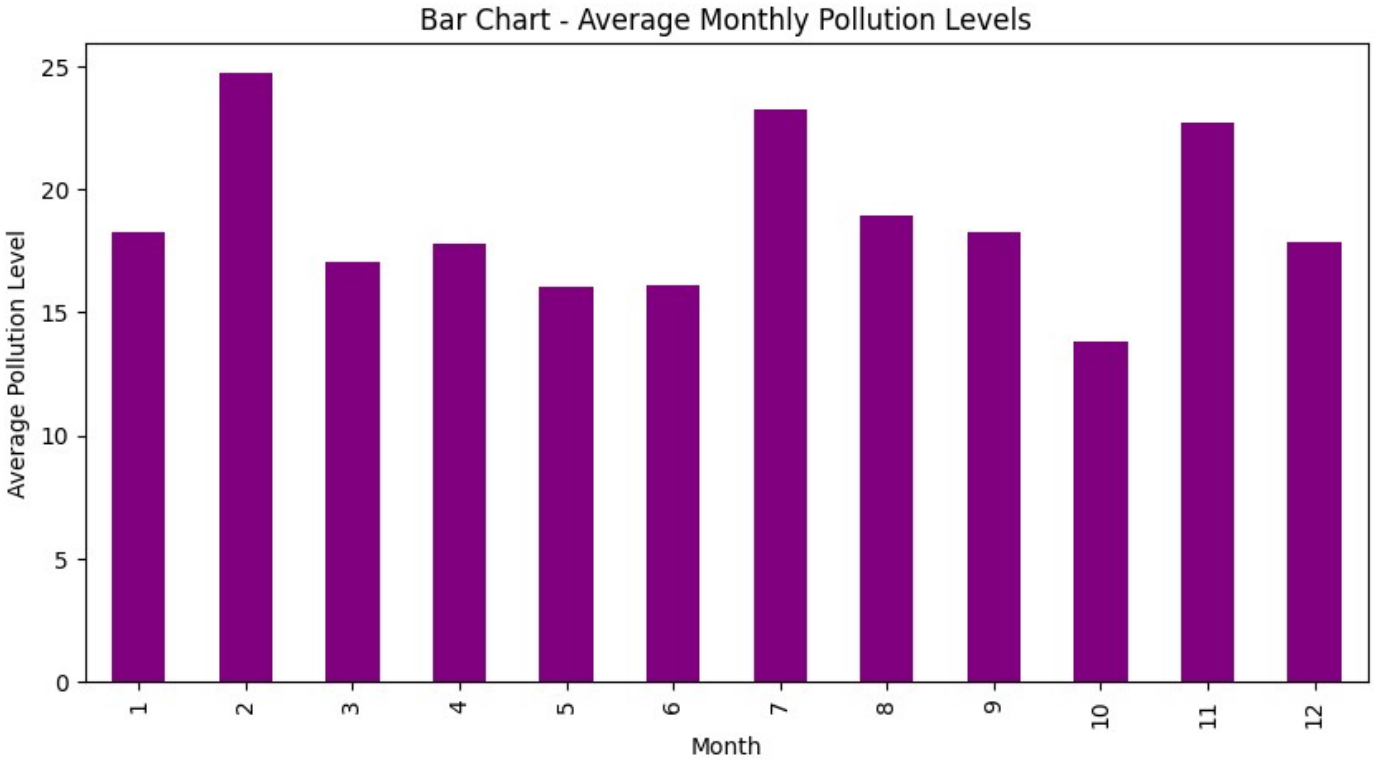
- Load and clean pollution data (2012-2021).
- Handle missing values and remove outliers.
- Visualize trends using **line plots, scatter plots, area charts, bar charts, and box plots**.
- Identify seasonal and yearly variations.
- Analyze and compare pollution levels over time.

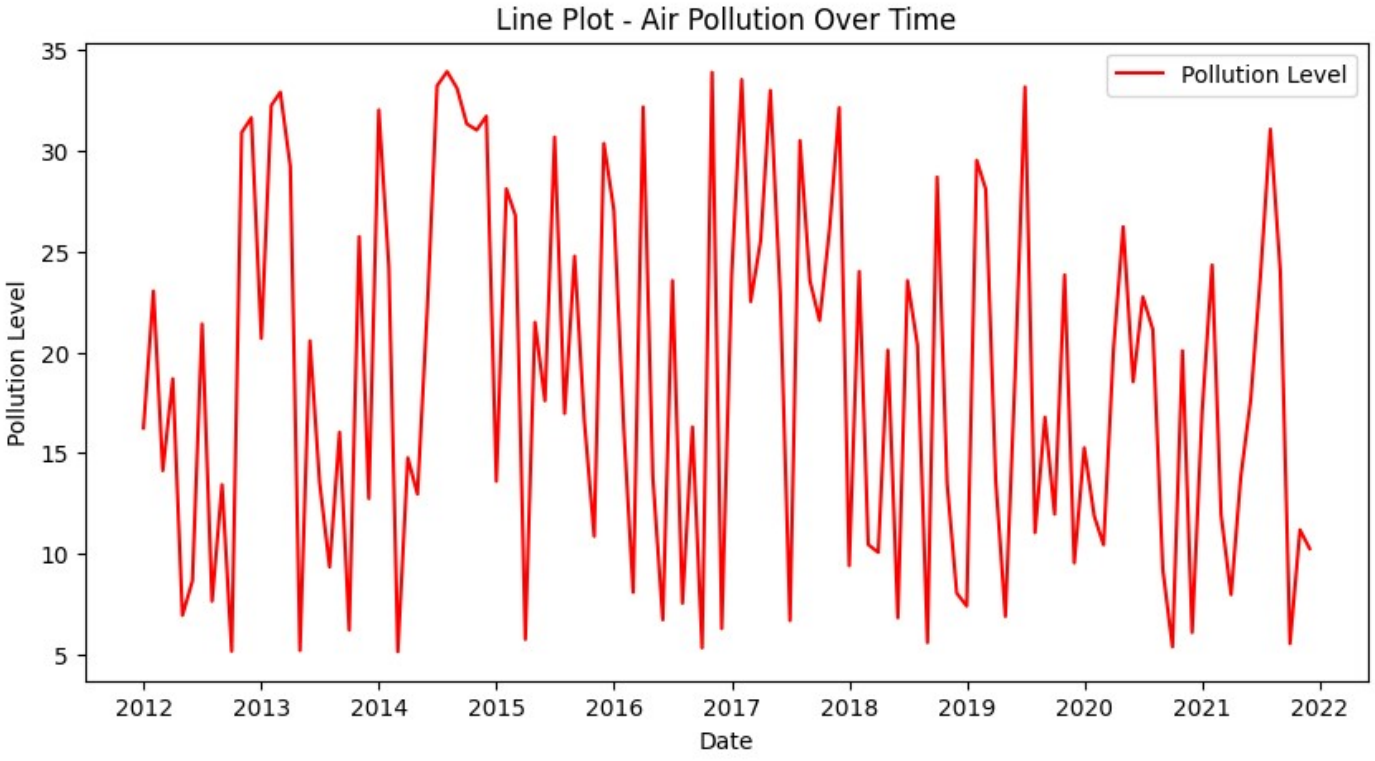
CODE:

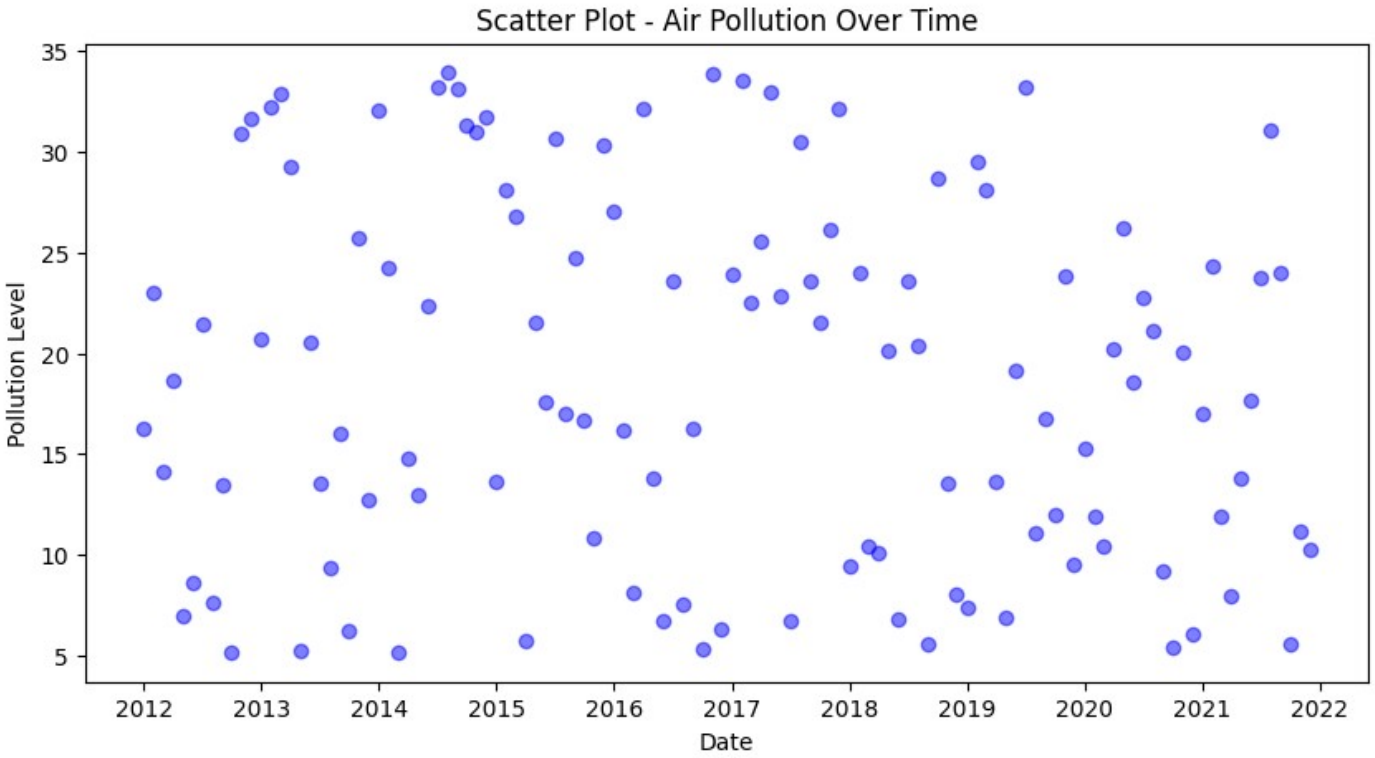
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the dataset
df = pd.read_csv("/content/us_Water_pollution_2012_2021_updated.csv")

# Convert 'Date' column to datetime format
df['Date'] = pd.to_datetime(df['Date'], errors='coerce')

# Select the pollution column (update the name if different)
pollution_col = "PM2.5 (µg/m³)"  # Update based on actual column name

# Filter data for 2012-2021
df = df[(df['Date'].dt.year >= 2012) & (df['Date'].dt.year <= 2021)]

# Remove outliers using IQR method
Q1 = df[pollution_col].quantile(0.25)
Q3 = df[pollution_col].quantile(0.75)
```

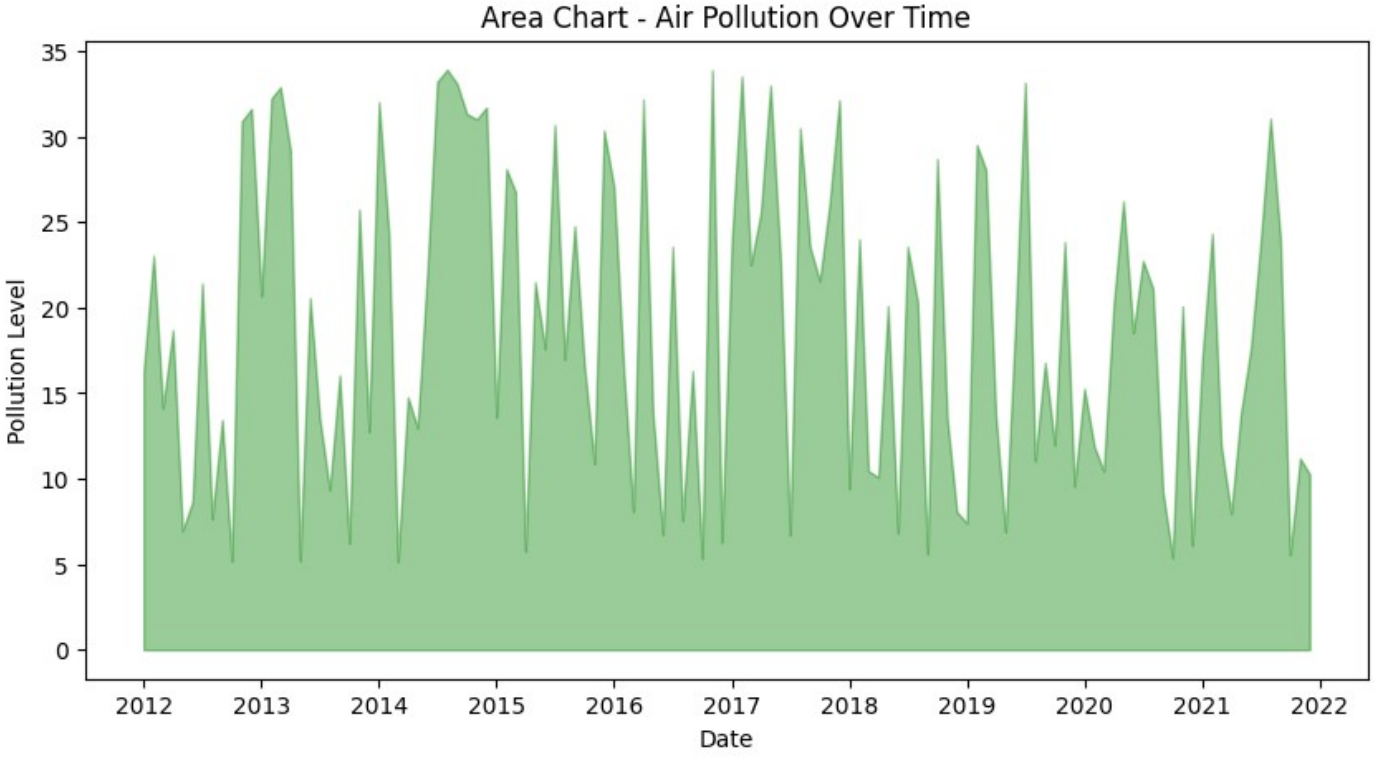
```
IQR = Q3 - Q1
df = df[(df[pollution_col] >= (Q1 - 1.5 * IQR)) & (df[pollution_col] <= (Q3 + 1.5 * IQR))]

# Set Date as index for plotting
df.set_index('Date', inplace=True)

# Plot 1: Line Plot
plt.figure(figsize=(10, 5))
plt.plot(df.index, df[pollution_col], color='red', label="Pollution Level")
plt.xlabel("Date")
plt.ylabel("Pollution Level")
plt.title("Line Plot - Air Pollution Over Time")
plt.legend()
plt.show()

# Plot 2: Scatter Plot
plt.figure(figsize=(10, 5))
plt.scatter(df.index, df[pollution_col], color='blue', alpha=0.5)
plt.xlabel("Date")
plt.ylabel("Pollution Level")
plt.title("Scatter Plot - Air Pollution Over Time")
plt.show()

# Plot 3: Area Chart
plt.figure(figsize=(10, 5))
plt.fill_between(df.index, df[pollution_col], color='green', alpha=0.4)
plt.xlabel("Date")
plt.ylabel("Pollution Level")
plt.title("Area Chart - Air Pollution Over Time")
plt.show()

# Plot 4: Bar Chart (Monthly Average)
df['Month'] = df.index.month
monthly_avg = df.groupby("Month")[pollution_col].mean()
monthly_avg.plot(kind="bar", color='purple', figsize=(10, 5))
plt.xlabel("Month")
plt.ylabel("Average Pollution Level")
plt.title("Bar Chart - Average Monthly Pollution Levels")
plt.show()

# Plot 5: Box Plot (Pollution Distribution)
plt.figure(figsize=(8, 5))
sns.boxplot(x=df.index.year, y=df[pollution_col], palette="coolwarm")
plt.xlabel("Year")
plt.ylabel("Pollution Level")
plt.title("Box Plot - Pollution Distribution by Year")
plt.xticks(rotation=45)
plt.show()
```

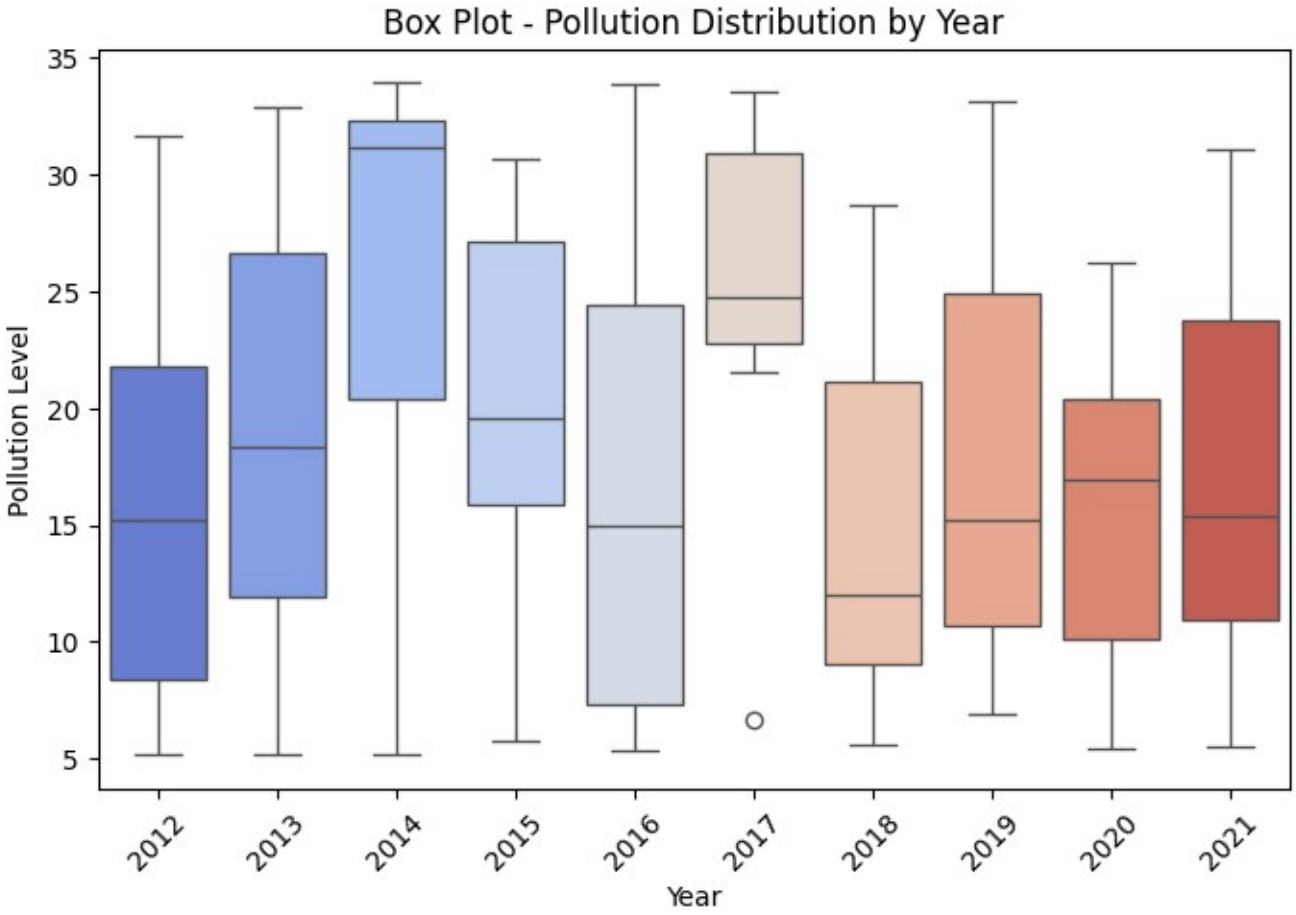
OUTPUT:


Bar Chart - Average Monthly Pollution Levels
Average Pollution Level
Month
0
5
10
15
20
25
1
2
3
4
5
6
7
8
9
10
11
12


Line Plot - Air Pollution Over Time
Pollution Level
Date
Pollution Level
5
10
15
20
25
30
35
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022


Scatter Plot - Air Pollution Over Time
Pollution Level
Date
35
30
25
20
15
10
5
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022


Area Chart - Air Pollution Over Time
Pollution Level
Date
35
30
25
20
15
10
5
0
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022



RESULT:

Thus, the program using the time series data implementation has been done successfully.